Project Description: Operation Analytics is the analysis done for the complete end to end operations of a company. With the help of this, the company then finds the areas on which it must improve upon. You work closely with the ops team, support team, marketing team, etc and help them derive insights out of the data they collect.

Being one of the most important parts of a company, this kind of analysis is further used to predict the overall growth or decline of a company’s fortune. It means better automation, better understanding between cross-functional teams, and more effective workflows.

Investigating metric spike is also an important part of operation analytics as being a Data Analyst you must be able to understand or make other teams understand questions like- Why is there a dip in daily engagement? Why have sales taken a dip? Etc. Questions like these must be answered daily and for that it’s very important to investigate metric spike.

Approach: I spent time on understanding the data and what all things I need to review. First, I created the database ‘operation analytics ‘and table using sql commands. Then I performed analysis to gain insights from the data.

Case Study 1 Operation Analytics:

A. **Number of jobs reviewed:** Amount of jobs reviewed over time.
Your task: Calculate the number of jobs reviewed per hour per day for November 2020?

```
select
count(distinct job_id)/(30*24) as num_jobs_reviewed
from job_data
where
ds between '2020-11-01' and '2020-11-30'
```

b. Throughput: It is the no. of events happening per second.
Your task: Let’s say the above metric is called throughput. Calculate 7 day rolling average of throughput? For throughput, do you prefer daily metric or 7-day rolling and why?

```
select ds, jobs_reviewed, avg(jobs_reviewed)over(order by ds rows between 6 preceding and current row) as throughput_7

from  (select ds, count(distinct job_id) as jobs_reviewed from job_data where ds between '2020-11-01' and '2020-11-30’

group by ds

order by ds)a
```

c. **Percentage share of each language:** Share of each language for different contents.
Your task: Calculate the percentage share of each language in the last 30 days?

option 1:

```
select language, num_jobs,
100.0* num_jobs/total_jobs as pct_share_jobs
from
(
select language, count(distinct job_id) as num_jobs
from job_data
group by language
)a
cross join
(
select count(distinct job_id) as total_jobs
from job_data
)b;
```

or
option 2:

```
select language,

num_language,total_language,

100.0* num_language/total_language as pct_share_lang

from

(

select language, count(language) as num_language

from job_data

group by language

)a

cross join

(

select count(language) as total_language

from job_data

)b;
```

d. **Duplicate rows:** Rows that have the same value present in them.
Your task: Let’s say you see some duplicate rows in the data. How will you display duplicates from the table?

Option 1:

```
select * from
(
select *,
row_number()over(partition by job_id) as rownum
from job_data
)a
where rownum>1
```

Option 2:

```
select job_id Cpunt(* ) as duplicates
```

```
from job_data
group by job_id
having count(*)>1
```

Case Study 2 Investigating metric spikes:

A. User Engagement: To measure the activeness of a user. Measuring if the user finds quality in a product/service.**Your task:** Calculate the weekly user engagement?

```
Select
extract (week from occurred_at) as weeknum,
Count(DISTINCT user_id)
from tutorial.yammer_events
group by weeknum
```

B. **User Growth:** Amount of users growing over time for a product.
Your task: Calculate the user growth for product?

```
select year, num_week, num_active_users,

sum(num_active_users) over(order by year, num_week rows between unbounded preceding and current row)

as cumm_active_users

from

(select

extract(year from activated_at) as year,

 extract(week from activated_at)as num_week,

count(distinct user_id) as num_active_users

from tutorial.yammer_users

where state='active'

group by year, num_week

order by year, num_week

)a
```

C. **Weekly Retention:** Users getting retained weekly after signing-up for a product.
Your task: Calculate the weekly retention of users-sign up cohort?

```
select count(user_id),
    sum(case when retention_week = 1 then 1 else 0 end) as per_week_retention
from
(
select a.user_id,
    a.sign_up_week,
    b.engagement_week,
    b.engagement_week - a.sign_up_week as retention_week
from
(
(select distinct user_id, extract(week from occured_at) as sign_up_week
from tutorial.yammer_events
where event_type = 'signup_flow'
and event_name = 'complete_signup'
and extract(week from occured_at)=18)a
left join
(select distinct user_id, extract(week from occured_at) as engagement_week
from tutorial.yammer_events
where event_type = 'engagement')b
on a.user_id = b.user_id
)
group by user_id
order by user_id;
```

D. **Weekly Engagement:** To measure the activeness of a user. Measuring if the user finds quality in a product/service weekly.
Your task: Calculate the weekly engagement per device?

```
select
extract(year from occurred_at) as year_num,
extract(week from occurred_at) as week_num,
device,
count(distinct user_id) as no_of_users
from tutorial.yammer_events
```

```
where event_type = 'engagement'
group by 1,2,3
order by 1,2,3;
```

E. **Email Engagement:** Users engaging with the email service.
Your task: Calculate the email engagement metrics?

```
select
100.0 * sum(case when email_cat = 'email_opened' then 1 else 0 end)
     /sum(case when email_cat = 'email_sent' then 1 else 0 end) as email_opening_rate,
100.0 * sum(case when email_cat = 'email_clicked' then 1 else 0 end) /sum(case when
       email_cat = 'email_sent' then 1 else 0 end)as email_clicking_rate
from
(
select *,
case when action in ('sent_weekly_digest', 'sent_reengagement_email')
  then 'email_sent'
  when action in ('email_open')
  then 'email_opened'
when action in ('email_clickthrough')
then 'email_clicked'
end as email_cat
from tutorial.yammer_emails
)a;
```

Tech Stack used: Mode for both creating database and finding insights.

Insights:

1) Case Study 1 Operation Analytics:

-The number of distinct jobs reviewed per hour per day for November 2020 is 83%.
- We used the 7-day rolling average of throughput as it gives the average for all the days right from day 1 to day 7 whereas

- The percentage share of Persian language is the most (37.5%).
-There are two duplicate rows if we partition the data by job_id.

Case Study 2 (Investigating metric spike):
- The weekly user engagement increased from week 18th to week 31st and then started declining from then onwards. This means that some of the users do not find much quality in the product/service in the last of the weeks.
• There are in total 9381 active users from 1st week of 2013 to the 35th week of 2014.
• The email opening rate is around 34% and email clicking rate is around 15%.

Results: With this detailed analysis we were able to gain insights into users
engagement and valuable insights related to users behaviour which will be used to predict the overall growth or decline of a company’s fortune.